Instituto Infnet

Projeto de Bloco

Data: 17/06/2022
Nome: Eduardo Marcello Duarte
Disciplina: Projeto em BI e Análise de Dados
Professor: Cassius Figueiredo

Introdução

O que seria um BI? Um BI se materializa através da importância de um dado ser estruturado na formação de uma informação, onde dará apoio na tomada de decisão de uma organização. Com essa técnica muitas empresas conseguiram aumentar seu valor no mercado usando métodos de análises para gerar novas estratégias, prever consequências, entre outras classificações.

Ao fazer esse trabalho eu pude notar os requisitos de um cientistas de dados enquanto passava pelas fases do trabalho, entendo sua importância e a necessidade de deles:

1. **Experiência ou habilidade de entender o negócio**: Habilidade de olhar aquele valor e ressaltar sua importância dentro da situação.
2. **Facilidade de atuar como profissionais de TI**: Necessário para localizar os dados, estruturas, ou modelos.
3. **Capacidade de aplicar matemática e estatística**: Não basta apenas entender os dados, todavia o cientista deve maximizar a capacidade de informação dele.

Com isso em mente eu os convido a lerem esse meu relatório, onde eu lhes apresentarei gráficos e tabelas para entendermos como foi esses nossos três anos de pandemia. Meu objetivo é trazer para vocês um texto simples e direto que mostrará todo o caminho que construí para responder uma simples pergunta: “Nesses três anos de pandemia, algum se destacou mais?”.

A partir das próximas páginas vocês entenderão de que fonte eu retirei minhas bases de dados, como eu as utilizei para gerar as informações necessárias para obter a resposta, e responderei não só a pergunta proposta como outras questões relacionadas a ela.

Esse meu projeto possuirá tabelas com os valores adquiridos através de cálculos e gráficos para nos dar uma questão mais visual para entendimento. Gostaria de ressaltar que tudo foi feito utilizando a linguagem de programação R na IDE do Rstudio, e que os códigos estarão disponíveis ao final do relatório.

Seguem o sumário a seguir para verem as etapas da leitura:

Sumário

Processamento,Limpeza e Transformação.

Aqui nesta etapa eu decidi pegar minhas bases de dados, retirados do site https://www.gov.br/saude/pt-br/composicao/se/demas/localizasus e construir uma base mais usável para os meus objetivos - as bases do site foram dadas em duas partes para os anos de 2020 e 2021. Devido a enorme quantidade de dados eu não conseguiria ter muita eficiência já que meu estudo está se baseando só no **município** do Rio de Janeiro então decidi coletar os dados necessários para construir a minha própria base com o foco naquilo que vou trabalhar.

Para a construção dessa nova base eu utilizei a linguagem de programação R na sua IDE RStudio. Foi bem simples a construção onde divide em 5 etapas para os anos de 2020 e 2021, o ano de 2022 por ter só uma parte não precisou da etapa 4 original. as etapas foram decididas em:

- 2020 e 2021:
    a. Criar variáveis para receber as duas partes da base;
    b. Através dessas duas variáveis terá uma filtração de dados usando a coluna **município (Rio de janeiro)**;
    c. Juntar as duas bases filtradas em uma só, checando se as colunas não foram resetadas em relação aos números em cada parte da base;
    d. Excluir as colunas que possuíam valores “NA”;
    e. Escrever um novo arquivo xlsx, ou csv, para salvar a nova base de dados.
- 2022:
    a. Cria uma variável para receber a base;
    b. Através dessa variável terá uma filtração de dados usando a coluna **município (Rio de janeiro)**;
    c. Excluir as colunas que possuíam valores “NA”;
    d. Escrever um novo arquivo xlsx, ou csv, para salvar a nova base de dados.

Com isso pude criar minha base de dados para consumo no projeto a qual estou desenvolvendo. Até agora meu foco se encontra nas colunas dos casos novos e óbitos onde irei comparar cada ano, no caso de 2022 compararei só os primeiros meses, entendendo a relação entre elas e futuramente adicionando a base de dados da vacina covid-19 para relacionar ela com os números de casos de cada ano.

Houve uma necessidade de alteração do tipo de dado de uma coluna. A coluna “data” dos três data frames (2020, 2021 e 2022) estava no formato “character” e portanto foi visto que deveria ser alterado para o tipo “Date”, o que ocorreu com êxito.

Durante a etapa de processamento foi feito uma análise básica em relação a cada coluna, usando o summary em cada data frame, aqui citarei as principais colunas que serão utilizadas durante o projeto,lembrando que as colunas são as mesmas nos três data frames (2020, 2021 e 2022):

- 2020
    - Data: Tipo character, mas transformado para date.
        - Min. : 2020-03-27
        - 1st Qu. : 2020-06-04
        - Median: 2020-08-13
        - Mean: 2020-08-13
        - 3rd Qu. : 2020-10-22
        - Max. : 2020-12-31
    - casosAcumulado: tipo numérico flutuante
        - Min. : 0
        - 1st Qu. : 34706
        - Median : 79183
        - Mean : 76392
        - 3rd Qu. : 115319
        - Max. : 165079
    - casosNovos: tipo numérico flutuante
        - Min. : 0.0

        - 1st Qu.: 4290
        - Median : 8782
        - Mean   : 7914
        - 3rd Qu.:11791
        - Max.   :14860
    - obitosAcumulado: tipo numérico flutuante
        - 1st Qu.: 4290
        - Median : 8782
        - Mean   : 7914
        - 3rd Qu.:11791
        - Max.   :14860
    - obitosNovos: tipo numérico flutuante
        - Min.   : -21.00
        - 1st Qu.: 17.75
        - Median : 44.50
        - Mean   : 53.07
        - 3rd Qu.: 78.25
        - Max.   :227.00
- 2021
    - Data:
        - Min.   :2021-01-01
        - 1st Qu.:2021-04-02
        - Median :2021-07-02
        - Mean   :2021-07-02
        - 3rd Qu.:2021-10-01
        - Max.   :2021-12-31
    - casosAcumulado:
        - Min.   :165158
        - 1st Qu.:229050
        - Median :369238
        - Mean   :354041

        - 3rd Qu.:484234
        - Max.   :500346
    - casosNovos:
        - Min.   :-2454.0
        - 1st Qu.: 196.0
        - Median : 570.0
        - Mean   : 918.5
        - 3rd Qu.: 1418.0
        - Max.   :12300.0
    - obitosAcumulado:
        - Min.   :14905
        - 1st Qu.:20791
        - Median :28805
        - Mean   :27326
        - 3rd Qu.:33982
        - Max.   :35190
    - obitosNovos:
        - Min.   : -1.0
        - 1st Qu.: 4.0
        - Median : 47.0
        - Mean   : 55.7
        - 3rd Qu.: 97.0
        - Max.   :246.0
- 2022
    - Data:
        - Min.   :2022-01-01
        - 1st Qu.:2022-01-24
        - Median :2022-02-16
        - Mean   :2022-02-16
        - 3rd Qu.:2022-03-11
        - Max.   :2022-04-03

- casosAcumulado:
  - Min.   :500346
  - 1st Qu.:677940
  - Median :891599
  - Mean   :809740
  - 3rd Qu.:929210
  - Max.   :954092
- casosNovos:
  - Min.   :   0
  - 1st Qu.: 752
  - Median : 1958
  - Mean   : 4879
  - 3rd Qu.: 6087
  - Max.   :23095
- obitosAcumulado:
  - Min.   :35190
  - 1st Qu.:35225
  - Median :35797
  - Mean   :35835
  - 3rd Qu.:36360
  - Max.   :36713
- obitosNovos:
  - Min.   :-2.00
  - 1st Qu.: 1.00
  - Median :12.00
  - Mean   :16.38
  - 3rd Qu.:24.00
  - Max.   :75.00

Análises

Porcentagem e total de casos por ano

Contaminação

Para a primeira análise foi utilizado as colunas “data”, “casosNovos”, e “casosAcumulado”. Primeiro foi pensado em achar total de dias dos três data frames então foi usado a função **difftime** para criar um data frame dos dias do ano, colocando o 2021-01-01 como limitador de 2020, 2022-01-01 como limitador de 2021 e 22-04-04 como limitador de 2022 (vai até o dia 22-04-03), como resultado foi achado respectivamente:

- 2020 com 280 dias ;
- 2021 com 365 dias;
- 2022 com 93 dias ;
- total de 738 dias.

Seguindo agora para achar o total de casos de cada ano, somando para comparar com o total dos casos acumulados (casosAcumulado é uma variável acumulativa que continua desde a data mínima de 2020 até a final do data frame de 2022), para achar o total dos casos foi usado a função **sum(dataframe$casosNovos)** e obtendo os seguintes valores:

- 2020 com 165.079 casos;
- 2021 com 335.267 casos;
- 2022 com 453746 casos;
- Total de 954.092 casos.

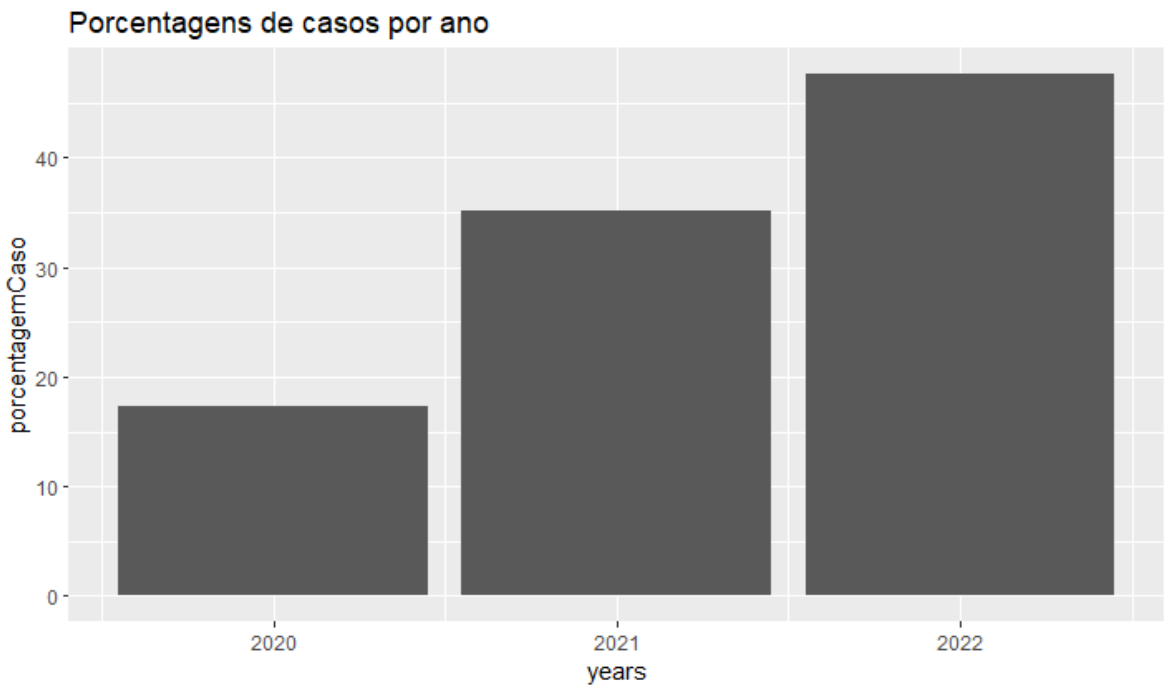
Após identificar o total de casos, foi necessário checar se ele batia com o último caso acumulado de 2022, que seria o absoluto dos casos, e foi um sucesso ao usar um if para fazer a comparação. Depois de ter o resultado foi necessário fazer os cálculos para adquirir as

porcentagens anuais, “(total de casos x 100) / total dos 3 anos” arredondando os resultados para ponto flutuante de uma casa decimal:

- 2020 com um percentual de 17.3%
- 2021 com um percentual de 35.1%
- 2022 com um percentual de 47.6%

Com isso pude obter todos os valores necessários para construir o primeiro gráfico e fazer a primeira análise que serve como ponto de partida para outras. Através de um data frame gerado com as colunas: “ano”, “dias”, ”casosAcumulados”, “porcentagemCaso”. Cada coluna irá receber os valores obtidos anteriormente e com esse data frame gerado e pronto para ser utilizado o gráfico foi montado usando o ggplot.

Year	Days	casosAcumulados	porcentagemCaso
2020	280	165.079,00	17,3
2021	365	335.267,00	35,1
2022	93	453.746,00	47,6



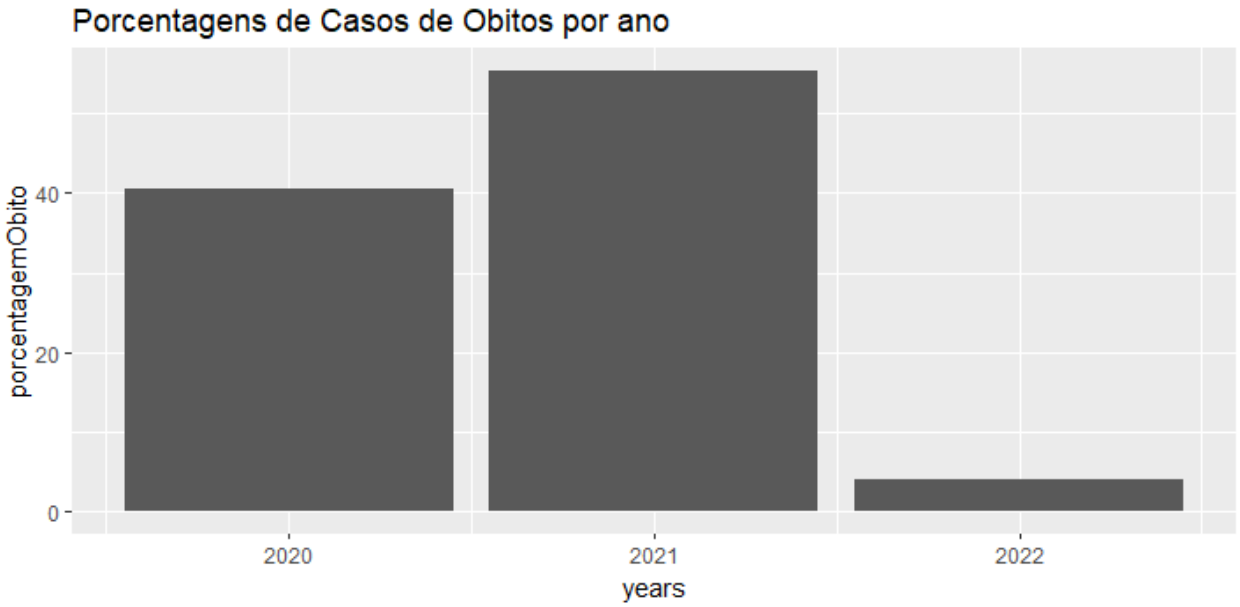
Com esse gráfico podemos ver que mesmo tendo apenas poucos dias com relação aos primeiros anos, o ano de 2022 possui uma grande porcentagem dos casos do COVID-19, muito provável ter tido relação com o surto da omicron no mês de janeiro onde teve quebra de recorde nos casos. Essa grade crescente nos permite ter a ideia do quão rápido aumentaram os casos de infecção, os picos podem estar relacionados com o avanço das variantes, como no caso da omicron, por isso a disparada entre 2020 e 2021 foi praticamente o dobro.

Óbitos

Seguindo a mesma linha de raciocínio dos casos de contaminação, trocando as variáveis para “ObitosNovos” e "ObitosAcumulados", pude construir uma nova tabela cujo o foco era mostrar os casos de óbitos totais por ano.

year	Days	obitosAcumulados	porcentagem Óbito
2020	280	14.860,00	40,5
2021	365	20.330,00	55,4

2022	93	1.523,00	4,1



Como podemos observar na tabela e no gráfico, diferente dos casos de contaminação a taxa de óbitos decresceu significativamente em 2022 sendo o ano com o maior número de contaminações. Acredito que isso seja graças ao avanço da vacinação que começou em 2021 com o intuito de diminuir ao máximo os casos graves e óbitos. O ano de 2021 possui o maior valor para casos de óbitos, enquanto 2022 os casos de contaminação.

Total de casos por mês e porcentagem de letalidade

Antes eu lhes apresentei o total de casos por ano, agora eu vou mostrar esses valores totais adquiridos separados mensalmente em cada ano, para que possamos obter uma visão mais detalhada.

Primeiro eu tive que extrair os dados necessários das bases com o intuito de construir novas tabelas de informações, e isso foi possível através do seguinte trecho do código:

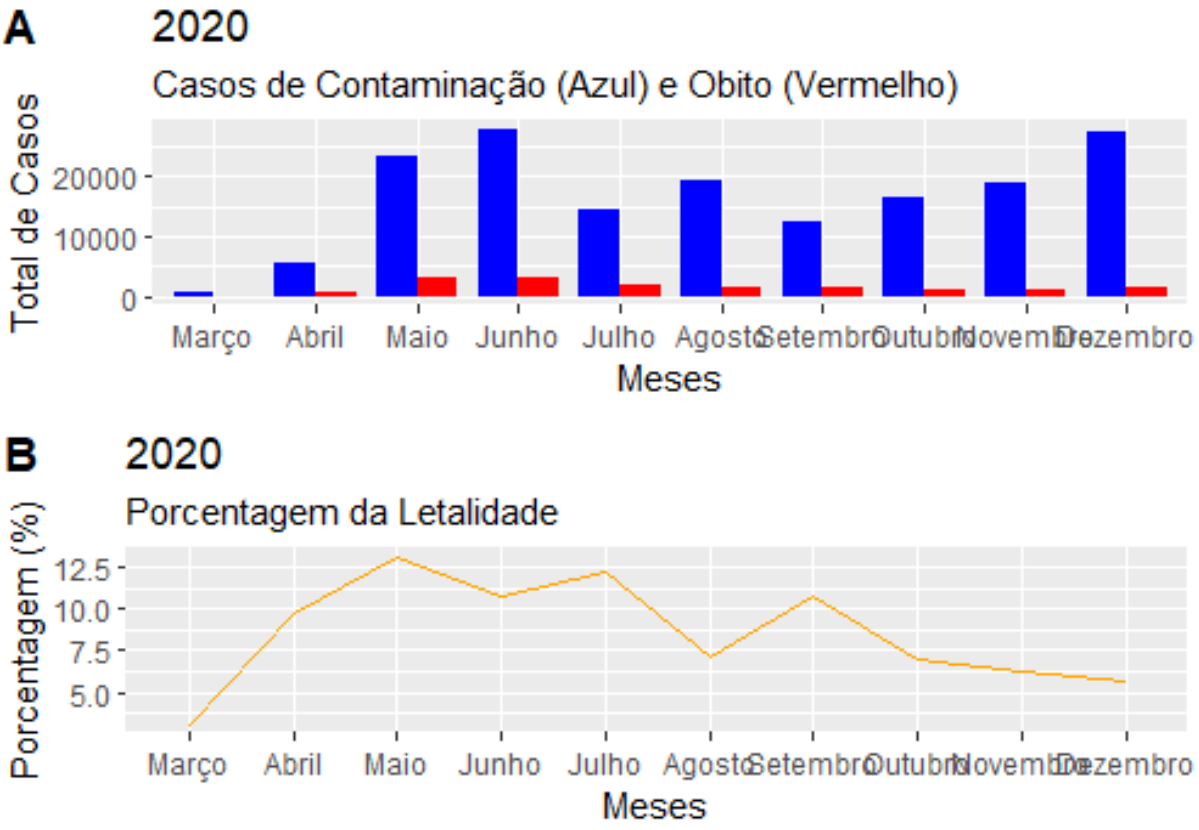
```
view_ano <- ano %>%
  ungroup() %>%
  mutate(Mes_Cod = format(data, "%m")) %>%
  group_by(Mes_Cod) %>%
        summarise(Casos = sum(casosNovos), Obitos = sum(obitosNovos))
```

Com as informações dos casos selecionados e separados pelo mês, eu fui atrás do cálculo da letalidade o que não foi difícil considerando a tabela (view_ano) montada, e só foi preciso construir uma coluna “Letalidade” e calcular o resultado, obtendo as tabelas desejadas. Agora mostrarei o trecho para calcular a letalidade e as tabelas adquiridas:

```
view_ano["Letalidade"]<-round((view_2020$Obitos/view_2020$Casos), 2)
```

2020

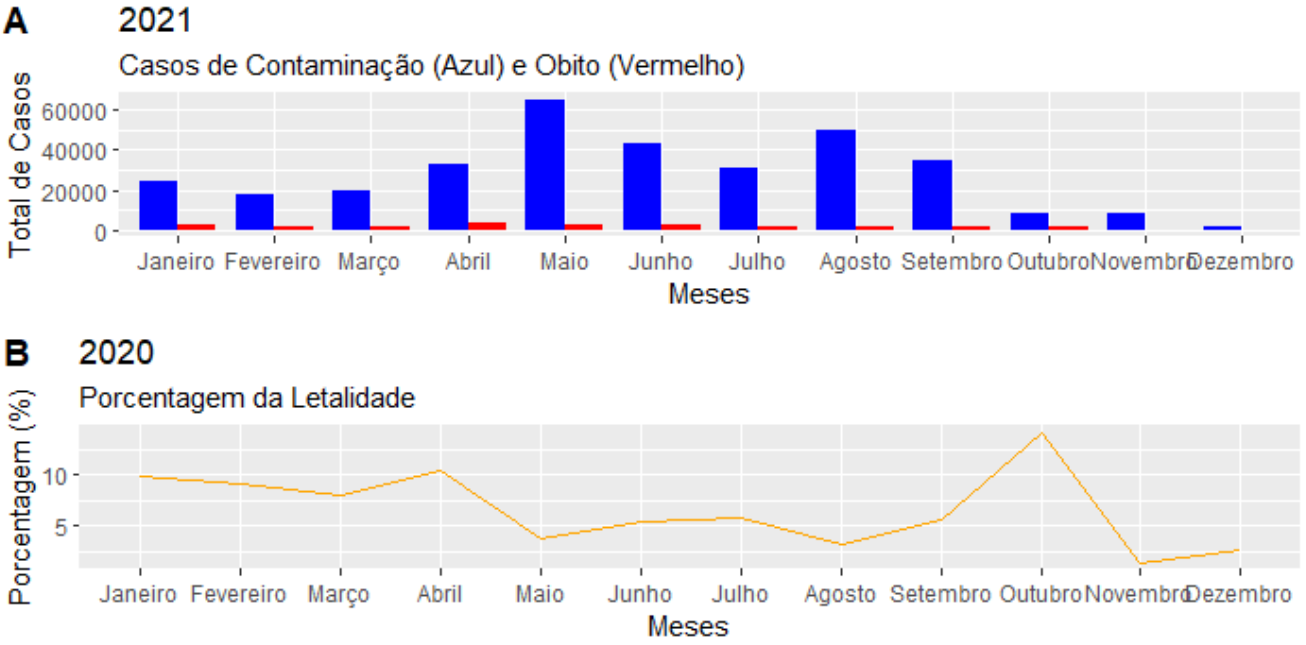
Mes	Mes_Cod	Casos	Obitos	Letalidade
Março	03	586,00	18,00	3,1
Abril	04	5.317,00	517,00	9,7
Maio	05	23.254,00	3.043,00	13,1
Junho	06	27.779,00	2.972,00	10,7
Julho	07	14.386,00	1.760,00	12,2
Agosto	08	19.144,00	1.353,00	7,1
Setembro	09	19.144,00	1.306,00	10,7
Outubro	10	16.485,00	1.156,00	7,0
Novembro	11	18.659,00	1.170,00	6,3
Dezembro	12	27.219,00	1.565,00	5,7



- Podemos ver que no primeiro ano de pandemia os meses em que ocorreram picos de contaminação foram: maio, junho e dezembro.
- Picos de óbitos se concentraram em maio e junho.
- Os meses com valores mais elevados de letalidade foram maio e julho.
- O mês de junho teve o maior valor para contaminação.
- O mês de maio teve os maiores valores para óbitos e porcentagem de letalidade.
- Podemos observar um crescimento na contaminação desde março até junho, com uma queda em julho, e novamente vindo a crescer do mês de setembro para dezembro.
- A taxa de óbitos só possui um crescimento contínuo de março até maio e depois novamente de setembro a dezembro.
- A letalidade teve um crescimento elevado de março para maio, mas após isso não obteve um valor que superasse o seu auge.

2021

Mes	Mes_Cod	Casos	Obitos	Letalidade
Janeiro	01	23.920,00	2.361,00	9,5
Fevereiro	02	17.986,00	1.642,00	9,1
Março	03	19.684,00	1.578,00	8,0
Abril	04	32.712,00	3.414,00	10,4
Maio	05	65.164,00	2.445,00	3,8
Junho	06	43.034,00	2.316,00	5,4
Julho	07	31.257,00	1.813,00	5,8
Agosto	08	49.666,00	1.571,00	3,2
Setembro	09	34.629,00	1.919,00	5,5
Outubro	10	7.951,00	1.124,00	14,1
Novembro	11	7.809,00	108,00	1,4
Dezembro	12	1.455,00	39,00	2,7

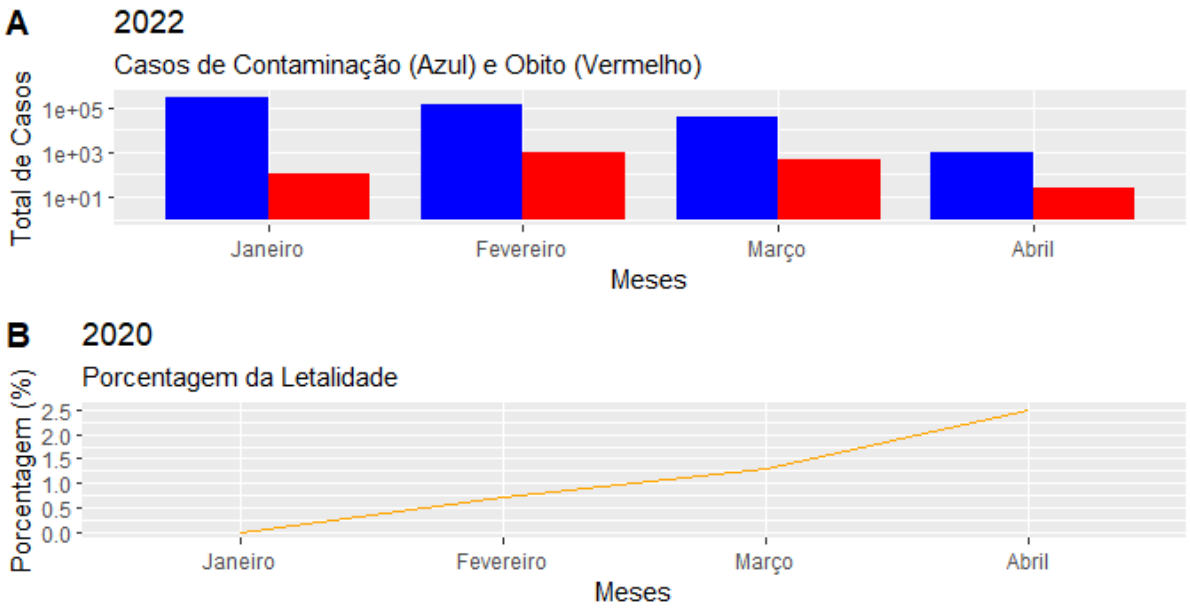


- No ano de 2021 os picos ocorrem em maio, junho e agosto.

- A taxa de óbitos possui seus picos nos meses de janeiro, abril, maio e junho.
- A porcentagem de letalidade possui valores altos nos meses de abril e outubro.
- O mês de maio possui o maior valor para casos de contaminação.
- O mês de abril possui o maior valor para os casos de óbitos.
- O mês de outubro possui a maior porcentagem de letalidade.
- Aqui o crescimento começa em março, após a queda de fevereiro em relação a janeiro, atingindo seu auge em maio.
- A taxa de óbitos possui mais quedas contínuas do que crescimentos.
- A letalidade teve mostrou ter um crescimento elevado do mês de setembro para outubro e consequentemente uma queda grande do auge para novembro.

2022

Mes	Mes_Cod	Casos	Obitos	Letalidade
Janeiro	01	284.909,00	101,00	0,0
Fevereiro	02	129.195,00	890,00	0,7
Março	03	38.724,00	509,00	1,3
Abril	04	918,00	23,00	2,5



- A taxa de contaminação possui picos nos meses de janeiro e fevereiro.
- A taxa de óbito possui valores altos em fevereiro e março.
- A porcentagem de letalidade possui valores altos em março e abril.
- Janeiro possui o maior valor da contaminação.
- Fevereiro possui o maior valor de óbitos.
- Abril possui a maior porcentagem de letalidade.
- A contaminação teve seu auge em janeiro, mas depois passou a cair nos próximos meses.
- A taxa de óbitos teve seu auge em fevereiro após um crescimento em relação a janeiro, todavia começa a cair como a taxa de contaminação.
- A porcentagem de letalidade possui um crescimento contínuo até atingir seu auge em abril.

Conclusão

Após obter todas as informações ditas anteriormente, pude me focar em obter a resposta que já foi dita durante a parte da introdução deste relatório. A pergunta “Nesses anos de pandemia, algum ano pode ser considerado o pior?” e venho a responder que não existe um ano pior, no pensamento geral, todavia cada ano pode possuir uma característica que se sobressai mais.

O ano de 2020 foi o começo da pandemia, então seus valores podem ser menores que seus anos posteriores, havendo a possibilidade dele até

possuir um valor maior que eles. Tirando o mês de março, os valores de contaminação e óbitos foram até altos, significando que mesmo sendo o ponto de partida, o COVID-19 conseguiu se proliferar bem.

2021 foi marcado mais pelas taxas de óbitos, todavia suas taxas de contaminações e as porcentagens de letalidade não ficam para trás. Todos os valores azuis ficaram na casa dos milhares, enquanto os vermelhos só tiveram dois valores nas casas das centenas e dezenas, que foi nos dois últimos meses.

2022 possui apenas quatro meses em sua base, sendo que abril não possui nem uma semana completa, mas foi o ano que mais teve contaminações ultrapassando os dois primeiros anos. Em relação à taxa de óbitos, foi o ano em que houve casos mínimos onde nenhum passou da casa das centenas e a letalidade não passou dos 3%.

Então com isso podemos concluir que:

1. Houve um ano que se sobressaiu mais que os outros?
   R: Não.

2. Por que?
   R: Como dito anteriormente, cada ano possui uma característica única que ultrapassa os outros.

3. O que marcou cada ano?
   R:
   - 2020 foi o começo da pandemia que se proliferou ao ponto dos valores de casos ultrapassarem a casa dos mil.
   - 2021 foi o auge dos casos de óbitos e a preservação das contaminações nas casas dos milhares.
   - 2022 foi marcado pelo pico de casos de contaminação onde quase chegou a ultrapassar os trezentos mil.

4. Podemos dizer que 2020 foi o menos pior?
   R: Não, pois mesmo possuindo valores menores de casos e letalidade, ele ainda foi o marco do início da pandemia tendo valores altíssimos.

5. Houve um aumento significativo de 2020 para 2021?
   R: Sim, ao revermos as tabelas e imagens postadas anteriormente veremos que houve sim um aumento significativo nos casos de contaminação e óbitos.

6. Houve alguma melhora de 2021 para 2022?
   R: Houve sim, mas em relação aos casos de óbitos e letalidade, as contaminações continuaram a crescer tendo um grande pico em janeiro.

7. Com as informações adquiridas podemos concluir que a vacina fez o seu efeito?
   R: Sim, pois o objetivo da vacina não era impedir que as contaminações diminuíssem, todavia minimizar ao máximo os casos graves e óbitos.

8. Com a vacinação é possível prever algum pico futuro?
   R: Observando os valores de óbitos de 2021 e 2022, acredita-se na dificuldade de picos que ultrapassam os valores de 2021.

Com isso eu encerro a proposta dita na introdução, espero que esse relatório esteja sendo do agrado de vocês tanto quanto agradou a mim, e que eu pude ajudar aqueles que queriam uma explicação mais simples sobre esses anos de pandemia que tivemos.Eu os agradeço por terem lido todo esse material feito por mim e para aqueles que quiserem ver os códigos completos, estarão nos anexos, após as referências.

Referências

https://covid.saude.gov.br

https://geekgeeky.com/pt-pt/como-alterar-a-escala-do-eixo-x-ou-y-em-r

https://r-graph-gallery.com/48-grouped-barplot-with-ggplot2.html

https://pt.stackoverflow.com/questions/394232/gráfico-de-barras-duplas

http://www.uel.br/projetos/experimental/pages/arquivos/GRAFICO_DE_BARRAS_2_FATORES.html

https://pt.stackoverflow.com/questions/124302/plotar-grafíco-barras-multiplas-variáveis-em-r

https://stackoverflow.com/questions/33221425/how-do-i-group-my-date-variable-into-month-year-in-r

Anexos

Código do Processamento

```
library(tidyverse)
library(xlsx)

# Achar a pasta do trabalho
if (!is.null(getwd())) setwd("D:/Users/Eduardo/Documents/BI/PB");

# Criar duas variáveis para receber as duas partes de 2020 da fonte escolhida
p1_20 <- read.csv("D:/Users/Eduardo/Documents/BI/PB/HIST_PAINEL_COVIDBR_2020_Parte1_03abr2022.csv", header = TRUE, sep = ";")
p2_20 <- read.csv("D:/Users/Eduardo/Documents/BI/PB/HIST_PAINEL_COVIDBR_2020_Parte2_03abr2022.csv", header = TRUE, sep = ";")

# Criar mais duas variáveis através do método de filtração das bases
rj_1 <- p1_20 %>% filter(municipio == "Rio de Janeiro")
rj_2 <- p2_20 %>% filter(municipio == "Rio de Janeiro")

# Juntar as duas bases em uma só
rj <- rbind(rj_1, rj_2)

# Excluir as colunas que possuem valores "NA"
rj$Recuperadosnovos <- NULL
rj$emAcompanhamentoNovos <- NULL

# Escrever um arquivo xlsx para salvar a base utilizável
write.xlsx(rj,file="COVIDRJ_2020.xlsx")

# Criar duas variáveis para receber as duas partes de 2021 da fonte escolhida
p1_21 <- read.csv("D:/Users/Eduardo/Documents/BI/PB/HIST_PAINEL_COVIDBR_2021_Parte1_03abr2022.csv", header = TRUE, sep = ";")
p2_21 <- read.csv("D:/Users/Eduardo/Documents/BI/PB/HIST_PAINEL_COVIDBR_2021_Parte2_03abr2022.csv", header = TRUE, sep = ";")
```

```
# Criar mais duas variáveis através do método de filtração das
bases
rj_1 <- p1_21 %>% filter(municipio == "Rio de Janeiro")
rj_2 <- p2_21 %>% filter(municipio == "Rio de Janeiro")

# Juntar as duas bases em uma só
rj <- rbind(rj_1, rj_2)

# Excluir as colunas que possuem valores "NA"
rj$Recuperadosnovos <- NULL
rj$emAcompanhamentoNovos <- NULL

# Escrever um arquivo xlsx para salvar a base utilizável
write.xlsx(rj,file="COVIDRJ_2021.xlsx")

# Criar uma variável para receber as parte de 2022 da fonte
escolhida
p_22 <-
read.csv("D:/Users/Eduardo/Documents/BI/PB/HIST_PAINEL_COVI
DBR_2022_Parte1_03abr2022.csv", header = TRUE, sep = ";")

# Criar mais uma variável através do método de filtração da base
rj <- p_22 %>% filter(municipio == "Rio de Janeiro")

# Excluir as colunas que possuem valores "NA"
rj$Recuperadosnovos <- NULL
rj$emAcompanhamentoNovos <- NULL

# Escrever um arquivo xlsx para salvar a base utilizável
write.xlsx(rj,file="COVIDRJ_2022.xlsx")
```

Código Transformação

```
library(tidyverse)
library(xlsx)
library(readxl)
library(lubridate)

if (!is.null(getwd())) setwd("D:/Users/Eduardo/Documents/BI/PB");

# Recebendo os dataframes.
```

```
p2020 <-
read.xlsx("D:/Users/Eduardo/Documents/BI/PB/COVIDRJ_2020.xlsx
", sheetName = "Sheet1")
p2021 <-
read.xlsx("D:/Users/Eduardo/Documents/BI/PB/COVIDRJ_2021.xlsx
", sheetName = "Sheet1")
p2022 <-
read.xlsx("D:/Users/Eduardo/Documents/BI/PB/COVIDRJ_2022.xlsx
", sheetName = "Sheet1")

# Checando a certeza dos dados.
sum(p2020$casosNovos)
sum(p2021$casosNovos + p2021$obitosNovos)

# Fazendo o summary de cada dataframe.
summary(p2020)
summary(p2021)
summary(p2022)

# Checando as colunas data para ver se estão no formato Date.
class(p2020$casosAcumulado)
class(p2021$data)
class(p2022$data)

# Transformando as colunas data de character para Date.
p2020$data <- as.Date(p2020$data)
p2021$data <- as.Date(p2021$data)
p2022$data <- as.Date(p2022$data)

# Checando os resultados.
class(p2020$data)
class(p2021$data)
class(p2022$data)
```

Código das análise anuais

```
library(tidyverse)
library(xlsx)
library(readxl)

if (!is.null(getwd())) setwd("D:/Users/Eduardo/Documents/BI/PB");

# Recebendo os dataframes.
```

```
p2020 <-
read.xlsx("D:/Users/Eduardo/Documents/BI/PB/COVIDRJ_2020.xlsx
", sheetName = "Sheet1")
p2021 <-
read.xlsx("D:/Users/Eduardo/Documents/BI/PB/COVIDRJ_2021.xlsx
", sheetName = "Sheet1")
p2022 <-
read.xlsx("D:/Users/Eduardo/Documents/BI/PB/COVIDRJ_2022.xlsx
", sheetName = "Sheet1")

# Checar os total de dias de 2020 e 2021
dias20 <- as.data.frame(difftime(p2020$data, as.Date("2021-01-01"),
units = "days"))
total_dias2020 <- dias20[1,1]
total_dias2020

dias21 <- as.data.frame(difftime(p2021$data, as.Date("2022-01-01"),
units = "days"))
total_dias2021 <- dias21[1,1]
total_dias2021

# 2022 vai da data 2022/01/01 até 2022/04/03
dias22 <- as.data.frame(difftime(p2022$data, as.Date("2022-04-04"),
units = "days"))
total_dias2022 <- dias22[1,1]
total_dias2022

# Achar o total de dias desses 3 anos
total_dias <- 280 + 365 + 93
total_dias

# Achar o total de casos por ano e comparar o somatório com o
último caso acumulado de 2022
total_casos2020 <- sum(p2020$casosNovos)
total_casos2020

total_casos2021 <- sum(p2021$casosNovos)
total_casos2021

total_casos2022 <- sum(p2022$casosNovos)
total_casos2022

1:nrow(p2022)
```

```
if((total_casos2020 + total_casos2021 + total_casos2022) ==
p2022$casosAcumulado[93]){
  # Total de casos acumulados nesses 738 dias é 954092
  TRUE
}

# A porcentagem dos casos de cada ano, levando em consideração
954092 como 100%
porc2020 <- round((total_casos2020 * 100 ) /
p2022$casosAcumulado[93], 1)
porc2020

porc2021 <- round((total_casos2021 * 100 ) /
p2022$casosAcumulado[93], 1)
porc2021

porc2022 <- round((total_casos2022 * 100 ) /
p2022$casosAcumulado[93], 1)
porc2022

if((porc2020 + porc2021 + porc2022) == 100){
  # Somatório das porcentagens é igual a 100
  TRUE
}

# Gerando um data frame juntando as respostas adquiridas
years <- c(2020, 2021, 2022)
view1 <- as.data.frame(years)
view1["Days"] <- c(280, 365, 93)
view1["casosAcumulado"] <- c(total_casos2020, total_casos2021,
total_casos2022)
view1["porcentagemCaso"] <- c(porc2020, porc2021, porc2022)

#Salvando a tabela view 1 em excel
write.xlsx(view1, "Casos_Contaminação_Anual.xlsx")

# Calculando o total de obitos por ano
total_obitos2020 <- sum(p2020$obitosNovos)
total_obitos2020

total_obitos2021 <- sum(p2021$obitosNovos)
total_obitos2021

total_obitos2022 <- sum(p2022$obitosNovos)
```

```
total_obitos2022

if((total_obitos2020 + total_obitos2021 + total_obitos2022) ==
p2022$obitosAcumulado[93]){
  TRUE
}

# Calculando a porcentagem de obitos por ano
porc2020 <- round((total_obitos2020 * 100 ) /
p2022$obitosAcumulado[93], 1)
porc2020

porc2021 <- round((total_obitos2021 * 100 ) /
p2022$obitosAcumulado[93], 1)
porc2021

porc2022 <- round((total_obitos2022 * 100 ) /
p2022$obitosAcumulado[93], 1)
porc2022

if((porc2020 + porc2021 + porc2022) == 100){
  # Somatório das porcentagens é igual a 100
  TRUE
}

# Gerando um data frame juntando as respostas adquiridas
years <- c(2020, 2021, 2022)
view2 <- as.data.frame(years)
view2["Days"] <- c(280, 365, 93)
view2["obitosAcumulado"] <- c(total_obitos2020, total_obitos2021,
total_obitos2022)
view2["porcentagemObito"] <- c(porc2020, porc2021, porc2022)

# Gerando o arquivo excel da tabela de obitos anuais
write.xlsx(view2, "Casos_Obito_Anual.xlsx")
```

Código dos cálculos mensais

```
library(tidyverse, warn.conflicts = FALSE)
library(xlsx)
library(readxl)

if (!is.null(getwd())) setwd("D:/Users/Eduardo/Documents/BI/PB");
```

```
# Recebendo os dataframes.
p2020 <-
read.xlsx("D:/Users/Eduardo/Documents/BI/PB/COVIDRJ_2020.xlsx
", sheetName = "Sheet1")
p2021 <-
read.xlsx("D:/Users/Eduardo/Documents/BI/PB/COVIDRJ_2021.xlsx
", sheetName = "Sheet1")
p2022 <-
read.xlsx("D:/Users/Eduardo/Documents/BI/PB/COVIDRJ_2022.xlsx
", sheetName = "Sheet1")

Mes <- c("Janeiro", "Fevereiro", "Março", "Abril", "Maio", "Junho",
"Julho",
         "Agosto", "Setembro", "Outubro", "Novembro", "Dezembro")

#Fazendo a view mensal de 2020
view_2020 <- p2020 %>%
  ungroup() %>%
  mutate(Mes_Cod = format(data, "%m")) %>%
  group_by(Mes_Cod) %>%
  summarise(Casos = sum(casosNovos), Obitos =
sum(obitosNovos))

view_2020["Mes"] <- Mes[3:12]

view_2020["Letalidade"] <- round((view_2020$Obitos /
view_2020$Casos), 2)

# Ordenando a tabela
order <- c("Mes", "Mes_Cod", "Casos", "Obitos", "Letalidade")
view_2020<- view_2020[, order]

#Fazendo a view mensal de 2021
view_2021 <- p2021 %>%
  ungroup() %>%
  mutate(Mes_Cod = format(data, "%m")) %>%
  group_by(Mes_Cod) %>%
  summarise(Casos = sum(casosNovos), Obitos =
sum(obitosNovos))

view_2021["Mes"] <- Mes

view_2021["Letalidade"] <- 0
```

```
view_2021$Letalidade <- round((view_2021$Obitos /
view_2021$Casos)*100, 1)

# Ordenando a tabela
order <- c("Mes", "Mes_Cod", "Casos", "Obitos", "Letalidade")
view_2021<- view_2021[, order]

#Fazendo a view mensal de 2022
view_2022 <- p2022 %>%
  ungroup() %>%
  mutate(Mes_Cod = format(data, "%m")) %>%
  group_by(Mes_Cod) %>%
  summarise(Casos = sum(casosNovos), Obitos =
sum(obitosNovos))

view_2022["Mes"] <- Mes[1:4]

view_2022["Letalidade"] <- 0
view_2022$Letalidade <- round((view_2022$Obitos /
view_2022$Casos), 2)

# Ordenando a tabela
order <- c("Mes", "Mes_Cod", "Casos", "Obitos", "Letalidade")
view_2022<- view_2022[, order]


write.xlsx(view_2020, "2020_Mensal.xlsx")
write.xlsx(view_2021, "2021_Mensal.xlsx")
write.xlsx(view_2022, "2022_Mensal.xlsx")
```

Código das visualizações dos gráficos

```
library(tidyverse)
library(xlsx)
library(readxl)
library(lubridate)
library(forcats)
library(ggpubr)

if (!is.null(getwd())) setwd("D:/Users/Eduardo/Documents/BI/PB");

# Recebendo as tabelas criadas nas outras etapas
```

```
view1 <-
read.xlsx("D:/Users/Eduardo/Documents/BI/PB/Casos_Contaminaçã
o_Anual.xlsx", sheetName = "Sheet1")
view2 <-
read.xlsx("D:/Users/Eduardo/Documents/BI/PB/Casos_Obito_Anual.
xlsx", sheetName = "Sheet1")
view_2020 <-
read.xlsx("D:/Users/Eduardo/Documents/BI/PB/2020_Mensal.xlsx",
sheetName = "Sheet1")
view_2021 <-
read.xlsx("D:/Users/Eduardo/Documents/BI/PB/2021_Mensal.xlsx",
sheetName = "Sheet1")
view_2022 <-
read.xlsx("D:/Users/Eduardo/Documents/BI/PB/2022_Mensal.xlsx",
sheetName = "Sheet1")

# Criando as Vizualizações dos calculos anuais em gráfico de barras
ggplot(view1, aes(years, porcentagemCaso)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Porcentagens de Casos de Contaminação por ano")

ggplot(view2, aes(years, porcentagemObito)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Porcentagens de Casos de Obitos por ano")

# Vizualizando os gráficos de casos mensais, por ano
meses_ord <- c("Janeiro", "Fevereiro", "Março", "Abril", "Maio",
"Junho", "Julho",
               "Agosto", "Setembro", "Outubro", "Novembro",
"Dezembro")

# 2020
p1 <- view_2020 %>%
  mutate(Meses = fct_relevel(Mes, meses_ord)) %>%
  ggplot(aes(Meses)) +
  geom_bar(aes(y = Casos),
           stat = "identity",
           fill = "blue",
           width = .4,
           position = position_nudge(x = -.20)) +
  geom_bar(aes(y = Obitos),
           stat = "identity",
           fill = "red",
           width = .4,
```

```
          position = position_nudge(x = .20)) +
  labs(
  title = "2020",
  subtitle = "Casos de Contaminação (Azul) e Obito (Vermelho)",
  x = "Meses",
  y = "Total de Casos",
  fill = "Casos"
  )

p2 <- view_2020 %>%
  mutate(Meses = fct_relevel(Mes, meses_ord)) %>%
  ggplot(aes(Meses)) +
  geom_line(aes(y = Letalidade),
            color = "orange",
            group = 1) +
  labs(
    title = "2020",
    subtitle = "Porcentagem da Letalidade",
    x = "Meses",
    y = "Porcentagem (%)",
    fill = "Letalidade"
  )

ggarrange(p1, p2,
          labels = c("A", "B"),
          ncol = 1, nrow = 2)

# 2021
p3 <- view_2021 %>%
  mutate(Meses = fct_relevel(Mes, meses_ord)) %>%
  ggplot(aes(Meses)) +
  geom_bar(aes(y = Casos),
           stat = "identity",
           fill = "blue",
           width = .4,
           position = position_nudge(x = -.20)) +
  geom_bar(aes(y = Obitos),
           stat = "identity",
           fill = "red",
           width = .4,
           position = position_nudge(x = .20)) +
  labs(
    title = "2021",
    subtitle = "Casos de Contaminação (Azul) e Obito (Vermelho)",
```

```
    x = "Meses",
    y = "Total de Casos",
    fill = "Casos")

p4 <- view_2021 %>%
  mutate(Meses = fct_relevel(Mes, meses_ord)) %>%
  ggplot(aes(Meses)) +
  geom_line(aes(y = Letalidade),
            color = "orange",
            group = 1) +
  labs(
    title = "2020",
    subtitle = "Porcentagem da Letalidade",
    x = "Meses",
    y = "Porcentagem (%)",
    fill = "Letalidade")

ggarrange(p3, p4,
          labels = c("A", "B"),
          ncol = 1, nrow = 2)

# 2022
p5 <- view_2022 %>%
  mutate(Meses = fct_relevel(Mes, meses_ord)) %>%
  ggplot(aes(Meses)) +
  geom_bar(aes(y = Casos),
           stat = "identity",
           fill = "blue",
           width = .4,
           position = position_nudge(x = -.20)) +
  geom_bar(aes(y = Obitos),
           stat = "identity",
           fill = "red",
           width = .4,
           position = position_nudge(x = .20)) +
  labs(
    title = "2022",
    subtitle = "Casos de Contaminação (Azul) e Obito (Vermelho)",
    x = "Meses",
    y = "Total de Casos",
    fill = "Casos") +
  scale_y_continuous(trans = "log100")

p6 <- view_2022 %>%
```

```
  mutate(Meses = fct_relevel(Mes, meses_ord)) %>%
  ggplot(aes(Meses)) +
  geom_line(aes(y = Letalidade),
            color = "orange",
            group = 1) +
  labs(
    title = "2020",
    subtitle = "Porcentagem da Letalidade",
    x = "Meses",
    y = "Porcentagem (%)",
    fill = "Letalidade")

ggarrange(p5, p6,
          labels = c("A", "B"),
          ncol = 1, nrow = 2)
```